

A Review on Finding Efficient Approach to Detect Customer Emotion Analysis using Deep Learning Analysis

Kottilingam Kottursamy

Associate Professor, Department of CSE, SRM Institute of Science and Technology, Kattankaluthur, Chennai, India.

E-mail: k.kottilingam@gmail.com

Abstract

The role of facial expression recognition in social science and human-computer interaction has received a lot of attention. Deep learning advancements have resulted in advances in this field, which go beyond human-level accuracy. This article discusses various common deep learning algorithms for emotion recognition, all while utilising the eXnet library for achieving improved accuracy. Memory and computation, on the other hand, have yet to be overcome. Overfitting is an issue with large models. One solution to this challenge is to reduce the generalization error. We employ a novel Convolutional Neural Network (CNN) named eXnet to construct a new CNN model utilising parallel feature extraction. The most recent eXnet (Expression Net) model improves on the previous model's inaccuracy while having many fewer parameters. Data augmentation techniques that have been in use for decades are being utilized with the generalized eXnet. It employs effective ways to reduce overfitting while maintaining overall size under control.

Keywords: Deep learning, sentiment analysis

1. Introduction

Since every household has at least one smart device these days, sensors are connected to them keep track of personal information. Gathering data like device usage habits and preferences isn't usually done with sensors but rather using mobile applications [1]. When the collected data is analysed and evaluated, it contains important information about the persons from whom it was taken. One of the primary goals of making predictions about consumers is to collect personal information. Emotional expressiveness is a crucial aspect of human intelligence [2]. The emotion classification is so essential for human-machine communication since it plays a far more significant role. Because expressions may be utilised in so many different ways in the real world, such as diagnosing autism, depression, retail, customer satisfaction, and education, classification of expressions may assist in improving each of these fields [3]. There is an increasing amount of researcher interest in emotion recognition, and several important articles have been published on the subject. All the applications discussed above rely on facial expressions acquired in a specific environment to generate an animated image on the fly. Because of this, FER in the wild generated a large amount of research and discussion [4]. Figure 1 shows the sample input image to find emotion from the customer in textile shop.



Figure 1. Sample input images of emotion detection

Our daily social and economic lives are increasingly integrated with intelligent and linked devices. Computers, smartphones, tablets, sensors, and cloud services have been utilised in both private and public places throughout the world in the last decades [5]. Data captured by these devices includes things like how much movement an individual does, such as walking, running, or climbing; how much sleep an individual receives; and where that individual goes [6]. Furthermore, new technology has provided new ways to link these devices to the human body, which means individuals no longer need to be physically there to receive information about them via the internet. For many individuals, dealing with everyday problems, such as stress and illness, is more difficult while they are online [7]. Managing the link between stress and mental health involves having emotional intelligence. Workers who perform critical tasks are often required to abide by regulations. Figure 2 shows basic emotion categories of customer's facial action.

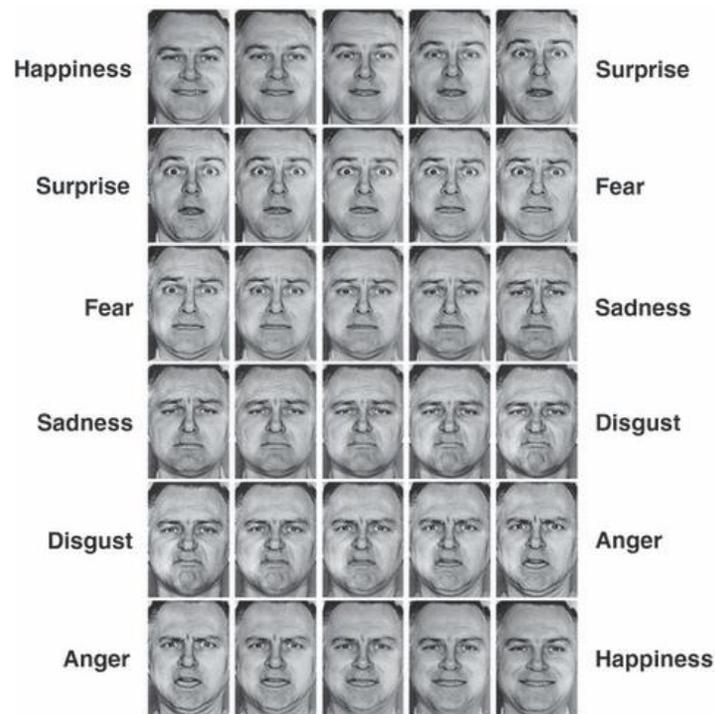


Figure 2. Sample Image Emotion Categories in the Dataset

For the most part, these significant developments came to fruition by using AlexNet and VGG, which used end-to-end techniques to classify an image according to its face expression [8, 9]. Adding additional layers and increasing the number of neurons per layer is a recurrent theme found in these projects. In order to get more accuracy, the number of layers has grown, but on the other hand, network sizes have increased, resulting in the overall width and height increasing (number of features per layer). The computational cost is proportional to the size of the network, so bigger networks tend to over fitting because of this [10]. Figure 3 shows the basic customer emotion categories.



Figure 3. Customer Emotion Categories

There have been several recent studies that claim to have developed state-of-the-art sentiment evaluation methods using deep learning algorithms. This study includes DNN, RNN, and CNN models including recurrent neural networks (RNN) and convolutional neural networks (CNN), which integrate convolutional long short-term memory (CNN-LSTM) to handle a range of sentiment analysis problems, such as sentiment polarity and aspect-based sentiment [11, 12]. Deep learning models were applied to datasets from Twitter, together with TF-IDF and word embedding, and we created state-of-the-art sentiment analysis techniques [13, 14].

2. Organization of the Research

This study paper includes several parts that demonstrate current understanding about consumer emotion detection analysis techniques, as section 3 describes. Section 4 details various state-of-the-art techniques for identifying consumer emotions, such as Deep Learning. Section 5 describes the deep learning benchmark datasets and their overall performance. The efficacy of the model is reviewed in section 6 and then at the end, the section covers customers' emotions.

3. Preliminaries

It has been discovered by a group of researchers, who are known as Glorot and his colleagues, that every facial expression is linked to an emotion [15]. For some human actions, such as emotions, body language, and speaking, the following facial expressions may be credited. Better emotion categorization was accomplished when combining facial features with audio information in both temporal and non-temporal modalities [16].

Many model named as VGG and GoogleNet are used as a foundation for the SHALLOW model of Arriaga et al which incorporates elements of both models. Fer-2013 is the dataset used to train their algorithm, and it results in a human-level performance rating of 65% [17, 18].

Additionally, Jain et al built a deep neural network with residual blocks for emotion classification. In this passage, the author argues that the network's accuracy is as up-to-date as possible by utilizing two popular facial datasets known as the CK+ and the Japanese Female Facial Expression (JAFFE). These datasets show accuracy levels of 93.24% and 95.23%, respectively. After the end of the first stage of two convolutional layers and a pooling layer, the final result was a combination of two residual blocks, which placed them on top of the others [19].

Liu et al used CNN for facial expression classification, thereby using CNN for the task of facial expression categorizing some experts have proposed three different types of networks, and each kind was shown to outperform its type on a wide range of emotional classes. The author arrived at a final FER-2013 precision of 65% by using a coordinated strategy on these three models [20].

In their most recent paper, Tautkute et al claim that they have the highest results on the CK+ and Indian Spontaneous Expression Database (ISED). Face expression classification was created using a deep neural network to identify landmarks in a person's face. To train the suggested network, AffecNet was used (a larger dataset from all previously available datasets for emotion classification). The cross-validation findings were further supported by the adoption of a fresh dataset that was more accurate [21].

Li et al conducted research and found that when utilizing a CNN, accuracy was only 70% [22]. This research by Sang et al had 71% accuracy on the same dataset, even though their model is not built on CNN. Using CNN and Support Vector Machine models, they were able to combine these two to get at this level of accuracy [23].

Agarwal et al note that the method for counting kernels and computing the size of the CNN utilized for emotion classification is: His two models included models with fixed and variable numbers of kernels and filters [24].

There are three different models based on CNN (Convolutional Neural Networks) which were recently introduced in a study by Shao et al. This network was designed for hardware constraints alone, which is one of the three training sets. While eXnet's shallow network only achieved 68% accuracy on the FER-2013 and 92.86% accuracy on the CK+ datasets, eXnet's frameworks [25].

In their study, Mehta et al. found that there was a correlation between how intense emotional states were and their intensity. Gradient histogram and local binary pattern are features that are manually configured. SVM was used to categorize different emotions [26].

4. Methodologies

In emotion classification, every technique, except the traditional methods, outperformed every other method. To do so, we developed CNN-based frameworks that operated at the same speed and used the same amount of resources regardless of the kind of operation. We offer the results of our thin network eXnet operating on an embedded system to show how well the performance is on real-time devices.

4.1 Naive Bayes Algorithm (NB)

A basic Bayesian classifier that employs frequent assumptions is used for classification tasks in many scenarios. The naive Bayes classifiers are often used since they are very popular among probabilistic machine learning models. Utilising Bayes' theorem, we arrive at conditional probabilities using mathematical proof [27]. Figure 4 shows basic structure of NB. The Bayes' theorem is used to determine the probability of "A" happening if "B" has occurred. This graphic depicts evidence "B" with a hypothesis "A". In the context of Bayesian model building, Naive Bayes believes that predictors and features are unrelated. To put it another way, the qualities are completely unrelated. Another fact is that it's also called for its naive. Naive Bayes is one of four different Naive Bayes techniques. One of the numerous practical uses of the Bayes theorem in computer science is to identify email as either "spam" or "not spam." In this study, we are reviewing the multinomial Naive Bayes technique.

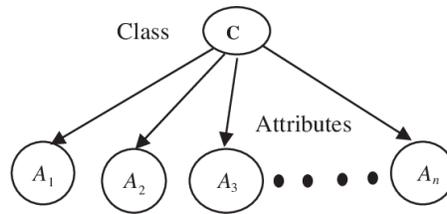


Figure 4. Basic Structure of NB

4.2 Support Vector Machine

The Support Vector Machine (SVM) was used as a foundation for the new Support Vector Regression technique. Linear regression does not account for margin, whereas SVR accounts for it. It cannot say that a place in space cannot be reached via the mainline and its total, because they are both capable of reaching any point in space. Minimize towards the effectiveness class,

$$= \frac{1}{2} \|W\|^2 + C \sum_{i=1}^N (\varepsilon_i + \varepsilon_i^*)$$

Condition 1:

$$y_i - wx_i - b \leq \varepsilon + e_i$$

$$wx_i + b - y_i \leq \varepsilon + e_i^*$$

$$\text{Where } e_i, e_i^* \geq 0$$

When describing the margin concept, the term "sum operation" is used. The location of the search vector completely depends on the dimensions of the space. The formula holds true in spaces with linear dimensions [28]. Because of the presence of eight variables, we live in an 8-dimensional space. The RBF kernel is used throughout the computation to utilise the radial basis function (RBF) kernel. Figure 5 shows error calculation in classification procedure.

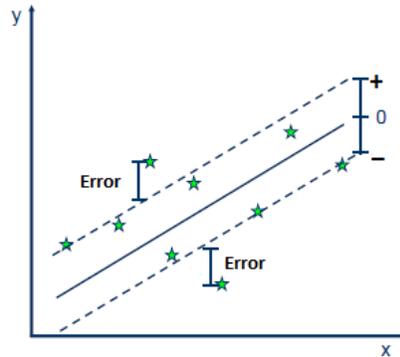


Figure 5. Error Calculation in Classification Process

4.3 Decision tree algorithm

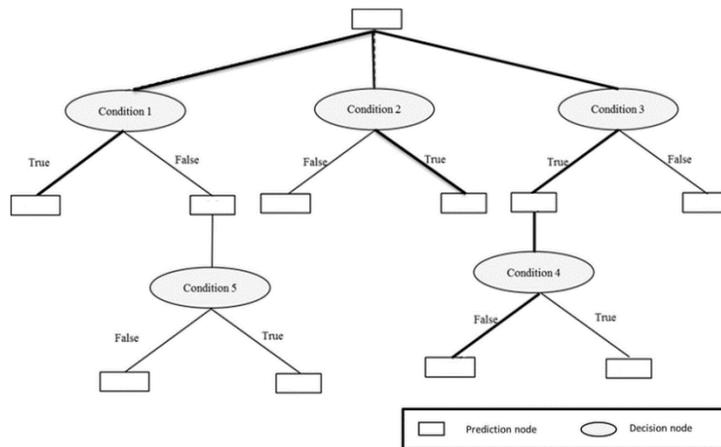


Figure 6 Basic Decision Tree Algorithm Structure

The Decision tree technique uses a tree-like graph to represent decisions and their possible consequences, including the outcomes of chance events, resource costs, and utility. A tree structure is used to calculate a classification function that predicts an attribute value given the values of the input attributes (variables). Classification is achieved using the Decision Tree, which takes labelled

datasets and breaks them into smaller datasets. According to Quinlan, this is the case. The final dataset will only include items that are highly linked after the dataset has been split into multiple datasets. In this process, the separated subsets are all connected to questions that have particular answers [29]. The assessment locates the subset of the new data that matches the criteria. Figure 6 shows basic tree structure of decision tree.

4.4 Random Forest

Random Forests and Random Decision Forests are used for classification and regression using supervised learning algorithms. Decision trees are known as "forests" when they have many tree instances. The training period is spent building a decision tree ensemble. Then, the ensemble's choices are combined to provide the target class (classification), which is decided by a majority vote of the decision trees or by the mean forecast of the individual trees (regression) [30]. Random Forest algorithms offer outstanding accuracy, and they are not vulnerable to overfitting, since the number of trees in the forest has an impact on their overall performance. Korting believes it (I'm paraphrasing, of course).

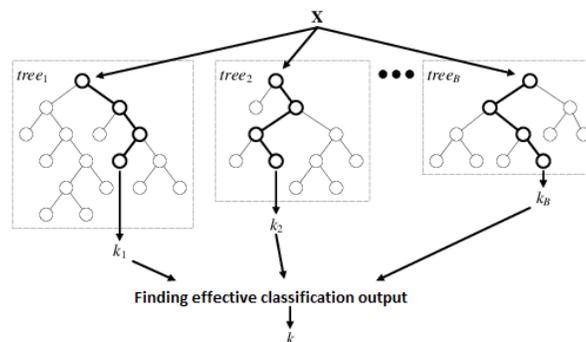


Figure 7. Effective Classification of RF

Bootstrapping training samples with decision trees yields a tree structure for a random forest. Instead of using the entire collection, decision tree splitting is performed randomly on a group of features and predictors. Ten estimators from the Random Forest algorithm are utilised in this research. Figure 7 shows tree structure of finding effective classification output in RF.

4.5 Convolution neural network (CNN)

With the current surge in deep learning through Convolutional Neural Networks CNN, many computer vision models have demonstrated excellent results. Because of their exceptional architecture, CNNs (55-layer deep learning nets) provide better performance than other machine learning methods. The extra hidden layers may be described as "feedforward neural networks," or "CNNs with hidden layers." CNN learns a representation from the data, by extracting features via hidden layers. Pooling and convolutional layers are often seen on many layers [31]. Figure 8 shows the basic block diagram of classification procedure by CNN.

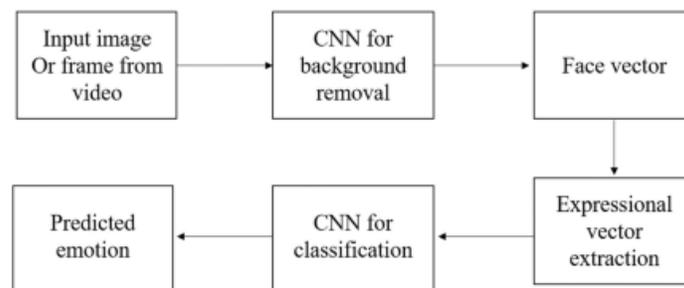


Figure 8. Basic Classification Procedure by CNN

Comparable to CNN, about 80% of the CNN image may be found in the convolutional layer. The data goes through several convolutional layers, each of which includes a filter (Kernels). Data is the raw material of the filters, and it is coupled with filtered data to generate a feature map that connects data with filtered data. When applying multiple filters to the input, a stack of feature

maps is produced, and this stack is the output of the convolutional layer. As a result, filter values are learned throughout the training process. You may get local information about the dependence or meaning when using the convolution operation. The technique used to implement nonlinearity in CNNs is the use of rectified linear units (RLUs) to construct feature maps. After the convolution process, a pooling layer is used to reduce the number of samples in each feature map and preserve the most important information.

The pooling layer shortens training time, while reducing overfitting. Pooling with max pooling occurs most often. CNN architectures are typically composed of convolutional layers interspersed with pooling layers that are further followed by many fully connected layers. The primary benefit of fast hardware performance is provided by a larger number of layers, as well as fewer parameters [32]. In fact, this serves to broaden network functionality within Occam's razor framework. Furthermore, we utilise many well-known data augmentation methods, all of which only require a single additional link in our network, in order to get the most performance out of our system.

4.5.1 Addition of Effective process

(a) Expression Network (eX-net)

Generally speaking, eXnet is divided into three stages of feature extraction. In the intermediate feature extraction phase, the features that will go on to the following step for extraction are extracted. But thanks to parallel extraction, our network has sparsity. The process is complete when you have arrived at the conclusion, known as the final step. Once you get to the conclusion, the process will finish. The eXnet design is better-suited for fast prototyping and early design cycles since it has fewer layers and parameters. The deep neural network for the emotion categorization process is a current trend for getting higher accuracy.

(b) Efficient activation function (EAF)

Convolution followed by max pooling and batch normalisation is done at this phase. Next comes batch normalisation, and followed by the Rectified Linear Unit (ReLU), which is also known as the Rectified Linear Unit (ReLU) (Convolution Batch-normalization ReLU). Our lightweight network encourages us to go the extra mile when organising the pooling process.

(c) Effective feature extraction (EFE)

We are suggesting the use of the Inception Net model as a model for parallel features and other feature extraction. To mimic how the brain processes data, in each instance it constructs two identical ParaFeat blocks. One of these blocks is set up to receive input serially, while the other is set up to process data alternately [33]. Route A goes over 1x1, 3x3, and identical convolutions; route B does the same. To minimise the number of parameters for both routes that use a single (1x1) convolution. While reducing parameters with the 1*1 convolution is only one of the many benefits of utilising this convolution filter, reducing parameters with this filter is shown by the 1*1 convolution. The fact that we are able to restrict the amount of parameters has enabled us to better generalise and make an efficient network simpler, enabling it to outperform on smaller datasets, such as this example. Concatenation of the route "A" feature map with the more distinctive route "B" feature map results in an improved representation of the input. A mixture of max pooling (2x2) and downsampling is used to downsample the image (2x2). To complete the overall eXnet architecture, this component performs image processing, including downsampling.

5. Results Discussion

At the ICML-2013 Challenges in Representation Learning, the most frequently used dataset for expression recognition was published, which contained over 35,000 gray-scale pictures of size (48*48) with seven fundamental emotions labelled. We used the ICML-2013 division,

which was given by the competition, to assess our approaches. Table 1 shows the efficacy rate improving by various deep learning algorithms.

Table 1 Efficiency Rate Computation of Various Algorithms

S.No	Algorithms	Accuracy	Classification Success rate	Improvement of efficiency	Improved Efficiency rate
1	Naïve Bayes	67.24%	72.45%	NO	0%
2	Single Classifier SVM	88.56%	91.89%	NO	0%
3	Decision Tree	83.89%	85.69%	NO	0%
4	Random forest	89.87%	90.87%	Partial	45%
5	CNN	94,34%	93.74%	Yes	80.45%
6	Proposed effective process in CNN	96.12%	98%	Yes	95.19%

An up-to-date Microsoft operating system with well recognised settings is used as the implementation environment for testing various techniques. Pytorch is used to build eXnet, a deep learning framework built using Pytorch.

$$\text{Efficiency of classification} = \frac{-1}{n} \sum_{i=1}^n p_i \ln(\hat{p}_i) + (1 - p_i) \ln(1 - \hat{p}_i)$$

Figure 9 shows the graph of overall performance measure for various algorithms. Cross-entropy loss is well-known, and it is done using a method that is well-known. Prior parametric parameters must be taken into account when training eXnet on the algorithm. After 60 epochs, the cyclical learning rate, which is originally set at FER-2013 and RAF-DB, begins to decline and continues to deteriorate until either the loss is rectified or 60 epochs have passed. Cross-validation is used to verify the CK+ dataset. The data augmentation method used on one dataset is applied to all datasets.

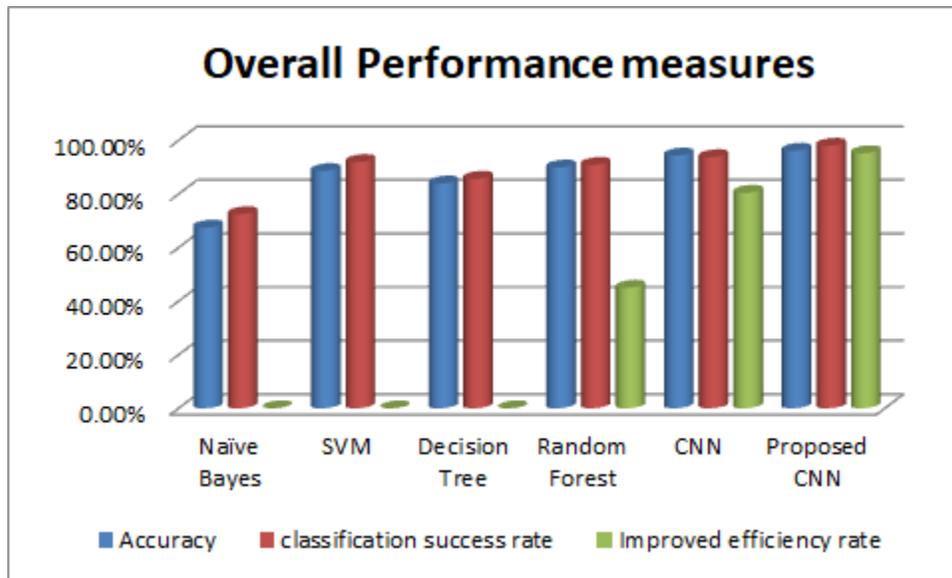


Figure 9. Overall Performance Measures for Various Algorithms

6. Conclusion

This research article has reviewed various deep learning algorithms and finding the possible fusion of effectiveness process of the algorithm. We will use the "quantization" method to significantly improve the overall effectiveness of eXnet in the wild as a future work. Even while this paradigm is very important, it has few limitations. The main difficulty in conducting this research was finding candidates who were willing to sign up for a one-year study. Just fifteen people from academia and business, both of whom were volunteering their time to agreed. It was difficult to provide training on the topics. It took participants a month to establish a habit of using the supplied application. Because of the small dataset size, the CPU is used in these experiments. It is a good solution to utilize GPUs, when the number of users increases. To supplement this study, we will look at how social media users feel about the studies we perform. We want to perform a Decision Integration Strategy (DIS) for deep learning machine learning techniques, as

well as a Decision Integration Strategy (DIS) for deep learning machine learning methods, to compare them to a conventional machine learning DIS.

References

- [1] Smys, S., and Wang Haoxiang. "Naïve Bayes and Entropy based Analysis and Classification of Humans and Chat Bots." *Journal of ISMAC* 3, no. 01 (2021): 40-49.
- [2] Li, S.; Deng,W. Reliable Crowdsourcing and Deep Locality-Preserving Learning for Unconstrained Facial Expression Recognition. *IEEE Trans. Image Process.* 2019, 28, 356–370.
- [3] Valanarasu, Mr R. "Comparative Analysis for Personality Prediction by Digital Footprints in Social Media." *Journal of Information Technology* 3, no. 02 (2021): 77-91.
- [4] Li, S.; Deng,W. Deep Facial Expression Recognition: A Survey. *arXiv* 2018, arXiv:1804.08348.
- [5] Manoharan, J. Samuel. "Capsule Network Algorithm for Performance Optimization of Text Classification." *Journal of Soft Computing Paradigm (JSCP)* 3, no. 01 (2021): 1-9.
- [6] Yang, H.; Han, J.; Min, K. A Multi-Column CNN Model for Emotion Recognition from EEG Signals. *Sensors* 2019, 19, 4736. doi:10.3390/s19214736.
- [7] Kumar, T. Senthil. "Construction of Hybrid Deep Learning Model for Predicting Children Behavior based on their Emotional Reaction." *Journal of Information Technology* 3, no. 01 (2021): 29-43.
- [8] Mehta, D.; Siddiqui, M.F.H.; Javaid, A.Y. Recognition of Emotion Intensities Using Machine Learning Algorithms: A Comparative Study. *Sensors* 2019, 19, 1897. doi:10.3390/s19081897.

- [9] Karunakaran, P., and Yasir Babiker Hamdan. "Early Prediction of Autism Spectrum Disorder by Computational Approaches to fMRI Analysis with Early Learning Technique." *Journal of Artificial Intelligence* 2, no. 04 (2020): 207-216.
- [10] Burkert, P.; Trier, F.; Afzal, M.Z.; Dengel, A.; Liwicki, M. DeXpression: Deep Convolutional Neural Network for Expression Recognition. arXiv 2015, arXiv:1509.05371.
- [11] Tripathi, Milan. "Analysis of Convolutional Neural Network based Image Classification Techniques." *Journal of Innovative Image Processing (JIIP)* 3, no. 02 (2021): 100-117.
- [12] Agrawal, A.; Mittal, N. Using CNN for facial expression recognition: A study of the effects of kernel size and number of filters on accuracy. *Visual Comput.* 2019, 2, 405–412.
- [13] Vijayakumar, T., and Mr R. Vinothkanna. "Efficient Energy Load Distribution Model using Modified Particle Swarm Optimization Algorithm." *Journal of Artificial Intelligence* 2, no. 04 (2020): 226-231.
- [14] Lopes, A.T.; de Aguiar, E.; Souza, A.F.D.; Oliveira-Santos, T. Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order. *Pattern Recognit.* 2017, 61, 610–628.
- [15] Sivaganesan, D. "Novel Influence Maximization Algorithm for Social Network Behavior Management." *Journal of ISMAC* 3, no. 01 (2021): 60-68.
- [16] Glorot, X.; Bordes, A.; Bengio, Y. Deep Sparse Rectifier Neural Networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, Lauderdale, FL, USA, 11–13 April 2011; pp. 315–323.
- [17] Suma, V. "Community Based Network Reconstruction for an Evolutionary Algorithm Framework." *Journal of Artificial Intelligence* 3, no. 01 (2021): 53-61.
- [18] Arriaga, O.; Valdenegro-Toro, M.; Plöger, P. Real-time Convolutional Neural Networks for Emotion and Gender Classification. arXiv 2017, arXiv:1710.07557.
- [19] Jain, D.K.; Shamsolmoali, P.; Sehdev, P. Extended deep neural network for facial emotion recognition. *Pattern Recognit. Lett.* 2019, 120, 69–74.

- [20] Liu, K.; Zhang, M.; Pan, Z. Facial Expression Recognition with CNN Ensemble. In Proceedings of the 2016 International Conference on Cyber worlds (CW), Chongqing, China, 28–30 September 2016; pp. 163–166.
- [21] Tautkute, I.; Trzcinski, T. Classifying and Visualizing Emotions with Emotional DAN. arXiv 2018, arXiv:1810.10529.
- [22] Li, S.; Deng, W. Deep Facial Expression Recognition: A Survey. arXiv 2018, arXiv:1804.08348.
- [23] Sang, D.V.; Van Dat, N.; Thuan, D.P. “Facial expression recognition using deep convolutional neural networks” In Proceedings of the 2017 9th International Conference on Knowledge and Systems Engineering (KSE), Hue, Vietnam, 19–21 October 2017; pp. 130–135. doi:10.1109/KSE.2017.8119447.
- [24] Agrawal, A.; Mittal, N. Using CNN for facial expression recognition: A study of the effects of kernel size and number of filters on accuracy. *Visual Comput.* 2019, 2, 405–412.
- [25] Shao, J.; Qian, Y. Three convolutional neural network models for facial expression recognition in the wild. *Neurocomputing* 2019, 355, 82–92.
- [26] Mehta, D.; Siddiqui, M.F.H.; Javaid, A.Y. Recognition of Emotion Intensities Using Machine Learning Algorithms: A Comparative Study. *Sensors* 2019, 19, 1897. doi:10.3390/s19081897.
- [27] Bindhu, V., and G. Ranganathan. "Effective Automatic Fault Detection in Transmission Lines by Hybrid Model of Authorization and Distance Calculation through Impedance Variation." *Journal of Electronics* 3, no. 01 (2021): 36-48.
- [28] Kulkarni, Tanay, Purnima Mokadam, Jnanesh Bhat, and Kailas Devadkar. "Potential customer classification in customer relationship management using fuzzy logic." In *International conference on innovative data communication technologies and application*, pp. 67-75. Springer, Cham, 2019.

- [29] Rasikannan, L., P. Alli, and E. Ramanujam. "Improved Feature Based Sentiment Analysis for Online Customer Reviews." In International Conference on Innovative Data Communication Technologies and Application, pp. 148-155. Springer, Cham, 2019.
- [30] Chacko, Anna Mariam, Bhuvanapalli Aditya Pranav, Bommanapalli Vijaya Madhvesh, and A. S. Poornima. "Customer Lookalike Modeling: A Study of Machine Learning Techniques for Customer Lookalike Modeling." In Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020, pp. 211-222. Springer Singapore, 2021.
- [31] Khedkar, Sujata, and Subhash Shinde. "Linguistic Feature-Based Praise or Complaint Classification from Customer Reviews." In International Conference on Intelligent Computing, Information and Control Systems, pp. 470-481. Springer, Cham, 2019.
- [32] Vargas, Jesus, Nelson Alberto, and Oswaldo Arevalo. "Algorithms for Decision Making Through Customer Classification." In Proceedings of International Conference on Intelligent Computing, Information and Control Systems, pp. 535-542. Springer, Singapore, 2021.
- [33] Mohan, JS Shyam, Hanumath Sreeman Vedantham, Venkata Chakradhar Vanam, and Nagendra Panini Challa. "Product Recommendation Systems Based on Customer Reviews Using Machine Learning Techniques." In Data Intelligence and Cognitive Informatics, pp. 267-286. Springer, Singapore, 2021.

Author's Biography

Kottilingam Kottursamy is currently working as an associate professor in the Department of CSE, SRM Institute of Science and Technology, Kattankaluthur, Chennai, India. His area of research includes bio statistics, machine learning algorithms, database tuning, high speed networks, information sciences & analytics.